



## Summary

### Main Features

#### PREAMBLE

Under the **Australian Bureau of Statistics Act 1975**, the Australian Statistician is responsible for formulating standards for the undertaking of operations for statistical purposes. The purpose of this paper is to provide information on the Australian Bureau of Statistics Data Quality Framework (ABS DQF). This framework provides the standards for assessing and reporting on the quality of statistical information. It is a tool which improves a user's ability to:

- decide whether a dataset or statistical product is fit for purpose (which in turn helps to identify data gaps);
- assess the data quality of seemingly similar collections; and
- interpret data.

It can also assist those developing statistical collections to produce high quality outputs.

The ABS Data Quality Framework is designed for use by a range of data users and providers in different settings, including government agencies, statistical agencies and independent research agencies. For example, the ABS DQF will be used to assess the quality of performance indicator data linked to a number of National Agreements in key policy areas signed by the Council of Australian Governments (COAG) in late 2008.

The ABS maintains ownership of this framework and reserves the right to update the framework as part of an ongoing commitment to continuous quality improvement.

#### INQUIRIES

For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070.

## Introduction

### INTRODUCTION

Among national statistical agencies, quality is generally accepted as "fitness for purpose". Fitness for purpose implies an assessment of an output, with specific reference to its intended objectives or aims. Quality is therefore a multidimensional concept which does not only include the accuracy of statistics, but also stretches to include other aspects such as relevance and interpretability.

Over the last decade, considerable work has been undertaken in statistical and economic agencies to define and measure quality. The ABS DQF is based on the **Statistics Canada Quality Assurance Framework** (2002) and the **European Statistics Code of Practice** (2005). The ABS DQF is comprised of seven dimensions of quality, reflecting a broad and inclusive approach to quality definition and assessment. The seven dimensions of quality are

**Institutional Environment, Relevance, Timeliness, Accuracy, Coherence, Interpretability and Accessibility.** All seven dimensions should be included for the purpose of quality assessment and reporting. However, the seven dimensions are not necessarily equally weighted, as the importance of each dimension may vary depending on the data source and context.

The ABS DQF has been designed to be used in evaluating the quality of statistical collections and products (e.g., survey data, statistical tables), including administrative data. Depending on the nature of the collection or product

being assessed, some dimensions will be more appropriate or important than others. For example, traditional measures of statistical accuracy for sample-based collections, such as sampling error and non-response error, may not apply to datasets which are by-products of administrative collections. For administrative data, other factors such as timeliness or relevance, for example, may be more important. We recommend that judgment is used in making assessments of quality, and that the quality dimensions are evaluated appropriately for the particular context.

This paper describes the ABS DQF, to enable its use in activities including the following:

- defining the quality of a data item or collection of data items (preparing a quality statement);
- assessing data in the context of a data need; and
- identifying data gaps and areas of future improvement.

Specifically, this paper overviews the ABS DQF, providing an explanation of each of the seven dimensions, followed by discussion to assist data users and producers to apply the framework. For each dimension, we state what constitutes the dimension, how it may be evaluated, and we suggest questions to be considered for the purpose of assessing the dimension.

## Institutional Environment

### INSTITUTIONAL ENVIRONMENT

The first dimension of quality in the ABS DQF is the Institutional Environment. This dimension refers to the institutional and organisational factors which may have a significant influence on the effectiveness and credibility of the agency producing the statistics. Consideration of the institutional environment associated with a statistical product is important as it enables an assessment of the surrounding context, which may influence the validity, reliability or appropriateness of the product.

**The dimension of Institutional Environment can be evaluated by considering six key aspects:**

- **Impartiality and objectivity:** whether the production and dissemination of data are undertaken in an objective, professional and transparent manner.
- **Professional independence:** the extent to which the agency producing statistics is independent from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, and potential conflict of interest.
- **Mandate for data collection:** the extent to which administrative organisations, businesses and households, and the public at large may be compelled by law to allow access to, or to provide data to, the agency producing statistics.
- **Adequacy of resources:** the extent to which the resources available to the agency are sufficient to meet its needs in terms of the production or collection of data.
- **Quality commitment:** the extent to which processes, staff and facilities are in place for ensuring the data produced are commensurate with their quality objectives.
- **Statistical confidentiality:** the extent to which the privacy of data providers (households, enterprises, administrations and other respondents), and the confidentiality of the information they provide, are guaranteed (if relevant).

The Institutional Environment dimension of a dataset or a statistical product can be evaluated by asking specific questions about the aspects listed above. We provide some suggestions of questions which might be asked, but these are not intended to be comprehensive or exhaustive. We encourage users and producers of statistics to generate their own questions to assess Institutional Environment in an appropriate way within their context.

**Suggested questions to assess Institutional Environment:**

- Which organisation(s) has supplied the data? What sort of organisation is this (e.g., public, commercial, non-government organisation)?
- Under what authority or legislation were the data collected?
- What procedures are in place to enable a need for a statistical product to be evaluated with respect to its scope, detail or cost?
- To what extent are quality guidelines documented by the agency?
- Is statistical confidentiality guaranteed, and if so, under what authority?
- To what extent, and how quickly, are any identified errors in published statistics corrected and publicised?

# Relevance

## RELEVANCE

The second dimension of quality in the ABS DQF is Relevance. This dimension refers to how well the statistical product or release meets the needs of users in terms of the concept(s) measured, and the population(s) represented. Consideration of the relevance associated with a statistical product is important as it enables an assessment of whether the product addresses the issues most important to policy-makers, researchers and to the broader Australian community.

**The dimension of Relevance can be evaluated by considering the following key aspects:**

- **Scope and coverage:** the purpose or aim for collecting the information, including identification of the target population, discussion of whom the data represent, who is excluded and whether there are any impacts or biases caused by exclusion of particular people, areas or groups.
- **Reference period:** this refers to the period for which the data were collected (e.g., the September-December quarter of the 2008-09 financial year), as well as whether there were any exceptions to the collection period (e.g., delays in receipt of data, changes to field collection processes due to natural disasters).
- **Geographic detail:** information about the level of geographical detail available for the data (e.g., postcode area, Statistical Local Area) and the actual geographic regions for which data are available.
- **Main outputs/ data items:** whether the data measures the concepts meant to be measured for its intended uses.
- **Classifications and statistical standards:** the extent to which the classifications and standards used reflect the target concepts to be measured or the population of interest.
- **Type of estimates available:** this refers to the nature of the statistics produced, which could be index numbers, trend estimates, seasonally adjusted data, or original unadjusted data.
- **Other cautions:** information about any other relevant issue or caution that should be exercised in the use of the data.

For more information about specific terms described above which are relevant to sample surveys (e.g., "scope", "coverage"), please see "[An Introduction to Sample Surveys: A User's Guide](#)".

To assist in evaluating the Relevance dimension of a dataset or a statistical product, we provide some suggestions of questions which might be asked below.

**Suggested questions to assess Relevance:**

- About whom, or what, were the data collected?
- Is there a time difference between the intended reference period, and the actual reference period of the collected data?
- How useful are these data at small levels of geography?
- Does this data source provide all the relevant items or variables of interest? Does the population presented by the data match the data need?
- To what extent does the method of data collection seem appropriate for the information being gathered?
- Have standard classifications (e.g., industry or occupation classifications) been used in the collection of the data? If not, why not?
- In what form are the statistics available? Are they original raw numbers, or indexes, or estimates?
- If rates and percentages have been calculated, are the numerators and denominators consistent?

# Timeliness

## TIMELINESS

Timeliness is the third dimension of quality in the ABS DQF. Timeliness refers to the delay between the reference period (to which the data pertain) and the date at which the data become available; and the delay between the advertised date and the date at which the data become available (i.e., the actual release date). These aspects are important considerations in assessing quality, as lengthy delays between the reference period and data availability, or between advertised and actual release dates, can have implications for the currency or reliability of the data.

**The dimension of Timeliness can be evaluated by considering two key aspects:**

- **Timing:** this refers to the time lag between the reference period and when the data actually become available (including the time lag between the advertised date for release and the actual date of release). For example, the reference period may be the 2004-05 financial year, but data may not become available for analysis until the middle of 2006.
- **Frequency of survey:** this refers to whether the survey or data collection was conducted on a one-off basis, or whether it is expected to be ongoing. If it is expected to be ongoing, frequency also includes information about the proposed frequency of repeated collections and when data will be released for subsequent reference periods.

To assist in evaluating the Timeliness dimension of a dataset or a statistical product, we provide some suggestions of questions which might be asked below.

**Suggested questions to assess Timeliness:**

- What is the gap of time between the reference period, the time when the data were actually collected, and the time when the data became available?
- Are there likely to be subsequent surveys or data collection issues for this topic?
- Are there likely to be updates or revisions to the data after official release?
- What is the gap between the advertised and actual release dates of the data?

## Accuracy

### ACCURACY

The fourth dimension of quality in the ABS DQF is Accuracy. Accuracy refers to the degree to which the data correctly describe the phenomenon they were designed to measure. This is an important component of quality as it relates to how well the data portray reality, which has clear implications for how useful and meaningful the data will be for interpretation or further analysis. In particular, when using administrative data, it is important to remember that statistical outputs for analysis are generally not the primary reason for the collection of the data.

Accuracy should be assessed in terms of the major sources of errors that potentially cause inaccuracy. Any factors which could impact on the validity of the information for users should be described in quality statements.

**The dimension of Accuracy can be evaluated by considering a number of key aspects:**

- **Coverage error:** this occurs when a unit in the sample is incorrectly excluded or included, or is duplicated in the sample (e.g., a field interviewer omits to interview a set of households or people in a household). Coverage of the statistical measures could be assessed by comparing the population included for the data collection to the target population.
- **Sample error:** where sampling is used, the impact of sample error can be assessed using information about the total sample size and the size of the sample in key output levels (e.g., number of sample units in a particular geographical area), the sampling error of the key measures, and the extent to which there are changes or deficiencies in the sample which could impact on accuracy.
- **Non-response error:** this refers to incomplete information provided by a respondent (e.g., when some data are missing, or the respondent has not answered all questions or provided all required information). Assessment should be based on non-response rates, or percentages of estimates imputed, and any statistical corrections or adjustment made to the estimates to address the bias from missing data.
- **Response error:** this refers to a type of error caused by respondents intentionally or accidentally providing inaccurate responses, or incomplete responses, during the provision of data. This occurs not only in statistical surveys, but also in administrative data collection where forms, or concepts on forms, are not well understood by respondents. Respondent errors are usually gauged by comparison with alternative sources of data and follow-up procedures.
- **Other sources of errors:** Any other serious accuracy problems with the statistics should be considered. These may include errors caused by incorrect processing of data (e.g. erroneous data entry or recognition), alterations made to the data to ensure the confidentiality of the respondents (e.g. by adding "noise" to the data), rounding errors involved during collection, processing or dissemination, and other quality assurance processes.
- **Revisions to data:** the extent to which the data are subject to revision or correction, in light of new information or following rectification of errors in processing or estimation, and the time frame in which

revisions are produced.

To assist in evaluating the Accuracy dimension of a dataset or a statistical product, we provide some suggestions of questions which might be asked below.

**Suggested questions to assess Accuracy:**

- Are there particular questions which are hard to understand and which respondents may provide an inaccurate response?
- To what extent are there procedures in place to manage processing error?
- Are any areas of the population unaccounted for in data collection?
- Are there particular questions which are sensitive and which respondents are less likely to answer?
- Have the data been adjusted in any way to account for non-response?
- Have the data been adjusted to ensure confidentiality of responses? If so, what methods have been used?
- What is the organisation's revision policy? How quickly are revisions produced and disseminated?
- Have the data been rounded at any stage in the collection or dissemination process?
- Has the sampling method changed for this data collection compared with previous cycles of data collection?
- Have weights been applied to the dataset? What are the benchmarks with which the weights align?

## Coherence

### COHERENCE

The fifth dimension of quality in the ABS DQF is Coherence. Coherence refers to the internal consistency of a statistical collection, product or release, as well as its comparability with other sources of information, within a broad analytical framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. Coherence is an important component of quality as it provides an indication of whether the dataset can be usefully compared with other sources to enable data compilation and comparison. It is important to note that coherence does not necessarily imply full numerical consistency, rather consistency in methods and collection standards. Quality statements of statistical measures must include a discussion of any factors which would affect the comparability of the data over time.

**The Coherence of a statistical collection, product or release can be evaluated by considering a number of key aspects:**

- **Changes to data items:** to what extent a long time series of particular data items might be available, or whether significant changes have occurred to the way that data are collected.
- **Comparison across data items:** this refers to the capacity to be able to make meaningful comparisons across multiple data items within the same collection. The ability to make comparisons may be affected if there have been significant changes in collection, processing or estimation methodology which might have occurred across multiple items within a collection.
- **Comparison with previous releases:** the extent to which there have been significant changes in collection, processing or estimation methodology in this release compared with previous releases, or any 'real world' events which have impacted on the data since the previous release.
- **Comparison with other products available:** this refers to whether there are any other data sources with which a particular series has been compared, and whether these two sources tell the same story. This aspect may also include identification of any other key data sources with which the data cannot be compared, and the reasons for this, such as differences in scope or definitions.

To assist in evaluating the Coherence dimension of a dataset or a statistical product, we provide some suggestions of questions which might be asked below.

**Suggested questions to assess Coherence:**

- Is it possible to compile a consistent time series of a particular data item of interest over a number of years?
- To what extent can a user meaningfully compare several data items within this collection?
- Could any natural disasters or significant economic events have influenced the data since the previous release?
- Have these data been confronted with other data sources, and are the messages consistent from all data sources?

# Interpretability

## INTERPRETABILITY

Interpretability is the sixth dimension of quality in the ABS DQF. Interpretability refers to the availability of information to help provide insight into the data. Information available which could assist interpretation may include the variables used, the availability of metadata, including concepts, classifications, and measures of accuracy. Interpretability is an important component of quality as it enables the information to be understood and utilised appropriately.

**The Interpretability of a statistical collection, product or release can be evaluated by considering two key aspects:**

- **Presentation of the information:** the form of presentation and the use of analytical summaries to help draw out the key message of the data
- **Availability of information regarding the data:** the availability of key material to support correct interpretation, such as concepts, sources and methods; manuals and user guides; and measures of accuracy of data.

To assist in evaluating the Interpretability dimension of a dataset or a statistical product, we provide some suggestions of questions which might be asked below.

**Suggested questions to assess Interpretability:**

- Are terms used in the statistical release or dataset which are ambiguous or likely to be confusing for a user?
- To what extent can a user of the release or dataset find supporting information about the data to enable improved interpretation?
- Are there information papers or articles available to help provide more insight into the concept(s) measured?
- Is there information available to help the user gauge the potential magnitude of error in the data?

# Accessibility

## ACCESSIBILITY

Accessibility is the seventh and final dimension of quality in the ABS DQF. Accessibility refers to the ease of access to data by users, including the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which information can be accessed. The cost of the information may also represent an aspect of accessibility for some users. Accessibility is a key component of quality as it relates directly to the capacity of users to identify the availability of relevant information, and then to access it in a convenient and suitable manner.

**The Accessibility of a statistical collection, product or release can be evaluated by considering two key aspects:**

- **Accessibility to the public:** the extent to which the data are publicly available, or the level of access restrictions. Additionally, special data services may include the availability of special or non-standard groupings of data items or outputs, if required.
- **Data products available:** this refers to the specific products available (e.g., publications, spreadsheets), the formats of these products, their cost, and the available data items which they contain.

To assist in evaluating the Accessibility dimension of a dataset or a statistical product, we provide some suggestions of questions which might be asked below.

**Suggested questions to assess Accessibility:**

- How easily can a user obtain this information? Is it publicly available?
- What range of products are available, and what are their costs?

- Are the data available in suitable formats?

## Applying the ABS Data Quality Framework

### APPLYING THE ABS DATA QUALITY FRAMEWORK

The ABS DQF is a general framework to enable a comprehensive and multi-dimensional assessment of the quality of a statistical dataset, product or release. It is intended that the framework enable data users and providers to:

- assess the quality of a data item or a collection of data items, with reference to the user's specific purpose and requirements; and
- design a statistical collection or product which is fit for purpose.

While ABS advises consideration of all seven quality dimensions, it is a matter of judgment as to the relative importance of each. We encourage users and producers to consider which quality dimensions are most relevant and important for their particular purpose. Quality relates to the fitness for purpose of the data or statistical product, and as purpose will vary among users, different users may make different assessments of the same product's quality. For example, if the credibility and trustworthiness of the data source are particularly important, then a careful examination of the Institutional Environment dimension will be especially important and this may have more weight in making an overall quality assessment. Alternatively, if a key purpose is to compare and contrast data, then the Coherence dimension will be particularly relevant.

#### Application of the ABS DQF by users of statistics

ABS recommends that when assessing the quality of a data item, dataset or other statistical product, a quality statement is developed. A quality statement is a presentation of information about the quality of a data item or a collection of data items, using the ABS DQF. The purpose of quality statements is to clearly communicate key characteristics of the data which impact on quality, so that potential users can make informed decisions about fitness for use. Quality statements should report both the strengths and limitations of the data.

Quality statements vary in length and detail, depending on the audience and medium for release. For example, the ABS has produced specific quality statements based on statistical releases called "quality declarations". Quality declarations are succinct summaries which quickly communicate key statistical quality messages, as well as providing links to more detailed information about statistical output. ABS quality declarations are designed primarily for electronic dissemination, hence their short length, and they enable layering of information in a web environment whereby each successive layer contains more detailed information. Quality declarations complement, but do not replace, the more comprehensive and complete ABS quality statements that currently exist (e.g., explanatory notes, and concepts, sources and methods documents).

#### Application of the ABS DQF by producers of statistics

The focus on the fitness of statistical information has emphasised the need to build quality into the production and delivery processes of collection agencies. The ABS recommends that producers of statistics consider the seven quality dimensions before designing collections, collecting statistics and producing outputs. This approach can enable informed decisions about factors including appropriate methodology, desired outputs and their accessibility, the coherence of the collection in relation to other collections or products and the relevance of the collection given its purposes.

Some suggested principles for managing each quality dimension are provided below.

#### ***Institutional environment***

Collection agencies should build a culture that focuses on quality, and an emphasise on objectivity and professionalism. Adequate resources and skills should be made available for the purpose intended. Cooperation of respondents can be encouraged by providing appropriate legal mandate and guarantees.

#### ***Relevance***

To be relevant, the collection agency must stay abreast of the information needs of its users. Mechanisms for

doing this include various consultative and intelligence-gathering processes, and regular stakeholder reviews.

### ***Timeliness***

The desired timeliness of the information derives from considerations of its main purposes: the period for which the information remain useful depends of the rate of change of the phenomenon being measured, the frequency of measure and the immediacy of the response that users may want to make based on the latest information. In addition to considering these aspects when planning target data release dates, consideration needs to be given to the capability of the organisation to produce the statistics within the given time frame. This capability includes staffing resources, system requirements, and the level of accuracy required of the data. The release of preliminary data followed by revised and final figures is often used a strategy for allowing less accurate data to be available sooner for decision making, with the subsequent release of more complete data occurring at a later stage.

### ***Accuracy***

Explicit consideration of the trade-offs between accuracy, cost and timeliness is important during the design stage. The coverage of the target population that can be achieved by the data collection strategy should be assessed. Proper testing of the instruments for data collection will ensure the reduction of response errors. Adequate measures have to be in place for encouraging response, following up non-response, and dealing with missing data (e.g., through imputation or adjustment made to the estimates). All stages of collection and processing should be subject to proper consideration of the need for quality assurance processes, including appropriate internal and external consistency checking of data with corresponding correction strategies.

### ***Coherence***

For managing coherence, collection agencies should use standard frameworks, concepts, variables and classifications, where such are available, to ensure the target of measurement is consistent over time and across different collections. As well, the use of common methodologies and systems for data collection and processing will contribute to coherence. Where data are available from different sources, consideration should be given to their confrontation and possible integration.

### ***Interpretability***

Managing interpretability is primarily concerned with the provision of sufficient information about the statistical measures and processes of data collection. Users need to know what has been measured, how it was measured and how well it was measured. The description of the methodology allows the user to assess whether the methods used were scientific or objective, and the degree of confidence they could have in the results. For meeting specific objectives, using analytical, descriptive or graphical techniques can often add value to help draw out the patterns in the data.

### ***Accessibility***

Management of accessibility needs to address how to help users know about the existence of the data or statistical product, locate it, and import it into their own working environment. Output catalogues, delivery systems, distribution channels and media, and strategies for engagement with users are all important considerations relating to this quality dimension.

## **MORE INFORMATION**

For more information on any of the issues discussed above please contact the Methodology and Data Management Division, ABS (Canberra) by email at [methodology@abs.gov.au](mailto:methodology@abs.gov.au), or by telephone via the ABS National Information and Referral Service on 1300 135 070.

## **About this Release**

This Information Paper describes the Australian Bureau of Statistics Data Quality Framework (ABS DQF), providing an explanation of each of the seven dimensions of the framework, followed by a discussion to assist data users and producers to apply the framework.

---



© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website - except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark - is licensed under a Creative Commons Attribution 2.5 Australia licence